

METODOLOGIE JĘZYKOZNAWSTWA

Ewolucja języka

Ewolucja teorii językoznawczych

pod redakcją

Piotra Stalmaszczyka

Wydawnictwo Uniwersytetu Łódzkiego

Łódź 2013¹

Zastosowanie korpusów językowych

w językoznawstwie kognitywnym

Małgorzata Fabiszak, Barbara Konat

Uniwersytet im. Adama Mickiewicza w Poznaniu

1. Pojęcie korpusowego językoznawstwa kognitywnego

(s. 131) W niniejszym artykule prezentujemy zarys metody badania też językoznawstwa kognitywnego za pomocą korpusów językowych. W pierwszej części artykułu zostanie zaprezentowane pojęcie korpusowego językoznawstwa kognitywnego, następnie przedstawimy szczegółowe omówienie etapów procesu badawczego w jednym z badań, aby na koniec opisać niektóre z korzyści i problemów związanych z omawianą metodą.

Korpusowe językoznawstwo kognitywne to połączenie założeń teoretycznych językoznawstwa kognitywnego oraz narzędzi badawczych językoznawstwa korpusowego (podobnie określa je na przykład Gries, 2007). Przez założenia teoretyczne w naukach rozumie się najogólniejsze podstawy determinujące kształt i zakres badań w danej dziedzinie (Nowak, 1977; Popper, 2002). Procedury badawcze z kolei, to różnorodne narzędzia pozwalające na sprawdzanie empiryczne hipotez wysnutych na podstawie pytań badawczych.

Założenia teoretyczne językoznawstwa kognitywnego w najogólniejszej postaci sformułował George Lakoff (1991). Przecistawiając kognitywizm lingwistyczny generatywizmowi, stwierdził on, że ten pierwszy określany jest przez dwa założenia czy też zobowiązania: założenie o generalizacji oraz założenie o poznawczym charakterze języka (ang. *generalization commitment* i *cognitive commitment*). W sformułowaniu Lakoffa założenie o generalizacji to „zobowiązanie do scharakteryzowania ogólnych zasad rządzących wszystkimi aspektami języka ludzkiego” (Lakoff, 1991, s. 53)². Wiąże się to z odrzuceniem pojęcia modularności umysłu i języka oraz z poszukiwaniem przez badaczy zależności o możliwie szerokim zakresie. Drugie założenie - założenie o poznawczym charakterze języka, zwraca (s.132) uwagę na mentalistyczny

¹ Artykuł opublikowany w bezpłatnym dostępie w Repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu za zgodą Wydawnictwa Uniwersytetu Łódzkiego. Całość publikacji dostępna na stronie Wydawnictwa: www.wydawnictwo.uni.lodz.pl.

² W oryginale: “The commitment to characterize the general principles governing all aspects of human language”.

(odziedziczony po generatywizmie) charakter językoznawstwa kognitywnego. Założenie to wprowadza tak zwany wymóg zbieżnego dowodu (ang. *converging evidence*): „aby koncepcje językoznawcze uzgadniać z tym, co generalnie wiadomo na temat umysłu i mózgu z dyscyplin innych niż językoznawstwo”³ (Lakoff, 1991, str. 54).

Językoznawstwo kognitywne jest dziedziną bardzo szeroką i zróżnicowaną. To konglomerat licznych teorii, czasem bardzo odmiennych od siebie. Do najważniejszych można zaliczyć: kognitywną teorię metafory (Lakoff i Johnson, 1999, 2003), teorię wyidealizowanych modeli kognitywnych (Lakoff, 1999), semantykę ram (Fillmore, 1975; Fillmore, 2006), teorię amalgamatu (Fauconnier, 1994), gramatykę kognitywną (Langacker, 1999), gramatykę konstrukcji (Goldberg, 2006), semantykę kognitywną (Taylor, 2007). Stosują one odmienny aparat pojęciowy, czasem też mówią o odmiennych przedmiotach badań, jednak jako wspólną podstawę przyjmują wyżej wymienione założenia.

Wśród bardziej szczegółowych założeń językoznawstwa kognitywnego wyróżniamy m. in. prymarność znaczenia i brak modularności języka w umyśle (Lewandowska-Tomaszczyk 2006, Geeraerts 2006, Evans, Green 2006). Te dwa założenia można prześledzić w trakcie badań nad polisemią, która jest jednym z podstawowych obszarów badawczych tego podejścia (np. Tuggy 1993, 1999, Lewandowska-Tomaszczyk 2002). Jeśli przyjąć, że językoznawstwo kognitywne polega na powiązaniu dynamicznych reprezentacji językowych z innymi procesami poznawczymi, w przypadku badań nad polisemią te inne zjawiska poznawcze o charakterze ogólnym to prototypowa struktura kategorii z elementami centralnymi i peryferyjnymi, odniesienie do schematów wyobrażeniowych i procesów poznawczych, np. skanowania sumarycznego (Langacker 1991) czy obiektywfikacji (Szwedek 2011).

Obok założeń językoznawstwa kognitywnego przyjmuje się, w omawianym nurcie, narzędzia i dane z korpusologii. Korpusy językowe to zbiory tekstów reprezentujące język mówiony i pisany, w określonych przez twórców korpusu proporcjach. Korpus może zawierać różne gatunki tekstów, np. teksty prasowe, fragmenty utworów literatury pięknej, teksty użytkowe, scenariusze, transkrypcje nagrań audycji radiowych, telewizyjnych, wykładów uniwersyteckich czy rozmów prywatnych. Często korpusy ogólne zawierają większą proporcję tekstów pisanych niż mówionych, ponieważ teksty takie nie wymagają pracochłonnego nagrywania i transkrypcji. Korpusy, których zawartość determinowana jest dostępnością tekstów raczej niż założeniami wypracowanymi przez ich twórców nazywamy korpusami oportunistycznymi. Dla niektórych lingwistów na miano korpusu tekstów zasługują tylko ustrukturyzowane bazy danych językowych, które poza samym tekstem zawierają informacje metatekstowe, takie jak gatunek (s.133) tekstu lub oznaczenie części mowy i ich cech gramatycznych. Takie otagowanie korpusu pozwala na przeszukiwanie go przy pomocy narzędzi komputerowych nie tylko pod względem poszczególnych

³ W oryginale: “The commitment to make one’s account of human language accord with what is generally known about the mind and brain from disciplines other than linguistics”. (Lakoff, 1991, str. 54)

leksemów, ale również form gramatycznych lub fraz wyrazowych. Jedne i drugie korpusy mogą służyć jako źródło przykładów zarówno do analiz jakościowych, jak i ilościowych. W badaniach ilościowych częstotliwość występowania i współwystępowania pewnych wyrażen i form można poddać analizie za pomocą różnych testów statystycznych, np. chi-kwadrat, standardyzacja Z, test studenta (Lewandowska-Tomaszczyk 2005), czy analiza korespondencji (mca) i regresja logistyczna (LogReg) (por. rozdziały w Glynn, Fischer 2010 oraz Glynn, Robinson w przygotowaniu).

Wskutek opisanego wyżej połączenia, w korpusowym językoznawstwie kognitywnym analizuje się interakcje pomiędzy możliwymi do zaobserwowania w korpusie zmiennymi. Zmienne te można podzielić najogólniej na semantyczne i formalne. Jak piszą autorzy z tego nurtu, w związku z przyjęciem niemodularnego charakteru wiedzy lingwistycznej przyjmuje się istnienie kontinuum leksykalno-syntaktycznego (Arppe, Gilquin, Glynn, Hilpert, Zeschel, 2010, s. 2). Dlatego też analizowanie struktur gramatycznych odbywa się w aparacie pojęciowym gramatyki kognitywnej (Langacker, 2009) czy konstrukcyjnej (Goldberg, 2006).

Jak zatem wygląda procedura badawcza w korpusowym językoznawstwie kognitywnym? Rozpoczyna się sformułowaniem pytania badawczego w aparacie pojęciowym językoznawstwa kognitywnego. Przy stawianiu pytań badawczych, językoznawcy używają pojęć metaforyzacji, konceptualizacji i konstruowania (Langacker 2009; Evans, Green 2006). Ten jakościowy opis badacz musi następnie sformułować w terminach ilościowych, dostępnych dla analiz korpusowych. Formułuje zatem przewidywania dotyczące relacji między zmiennymi w postaci hipotez, zawierających terminy ilościowe. Następnie testuje przewidywania wyrażone w hipotezach na dostępnym korpusie. W kolejnym kroku, badacz musi otrzymane wyniki przeformułować na jakościowy język opisu językoznawstwa kognitywnego.

2. Przykład zastosowania procedury badawczej

Spróbujmy zatem prześledzić działanie zaproponowanej powyżej procedury badawczej na przykładzie konkretnego badania z zakresu korpusowego językoznawstwa kognitywnego. Wybrane studium (Fabiszak, Hebda, Konat 2012) stawia sobie następujące pytanie badawcze: Jakie są różnice w językowym wyrażaniu doświadczeń mentalnych indywidualnych a kolektywnych (dostępnych intersubiektywnie)? Na poziomie założeń teoretycznych, pytanie to wyrasta z poszerzania się zainteresowań językoznawstwa kognitywnego. Poszerzanie to trafnie opisał Bernárdez (2007), który zauważa, że językoznawstwo kognitywne przeszło przez (s. 134) sześć etapów rozumienia semiozy – tworzenia znaczeń. W początkowym okresie skupiano się na roli mózgu i postrzegano rozumienie jako uwewnętrzniony proces mentalny zgodnie z metaforą POZNANIE TO MÓZG, co prowadziło do widzenia zjawisk językowych jako procesów autonomicznych (por. teoria Chomsky'ego). Wczesne językoznawstwo kognitywne (Lakoff i Johnson 1980, Lakoff 1987, Johnson 1987, Langacker 1991) podkreślało znaczenie ludzkiego ciała, człowieka jako organizmu biologicznego wchodzącego w interakcje ze środowiskiem w procesach poznawczych, co znalazło wyraz w hipotezie poznania ucieleśnionego (ang. *embodied cognition*). Badania nad

konceptualnymi metaforami prymarnymi (Grady 1997, Christopher Johnson 1999) i metaforami konceptualnymi o wyższym poziomie złożoności podkreśliły rolę kontekstu społeczno-kulturowego – poznanie usytuowane (ang. *situated cognition*) oraz wkład innych uczestników interakcji werbalnej – poznanie rozproszone (ang. *distributed cognition*) w tworzenie reprezentacji mentalnych (por. Sinha 2005, Zlatev et al. 2008). Wreszcie badania nad tworzeniem metafor w dyskursie (np. Nerlich, Hamilton, Rowe 2002, Cameron 2010, Semino 2011, Fabiszak, Hebda 2012) wpisują się w takie podejście do badania znaczenia, które uwzględnia również wymiar historyczny zjawisk językowych, a które Bernardez nazywa poznanem synergicznym (ang. *synergic cognition*).

Analizowany poniżej przykład badania z zakresu korpusowego językoznawstwa kognitywnego można postrzegać jako badanie nad poznanem rozproszonym, w którym analiza semantyczno-syntaktyczna ma wykazać różnice w konceptualizowaniu stanów mentalnych dwóch rodzajów – intersubiektywnie poznawalnych (poznanie kolektywne) oraz indywidualnych (poznanie indywidualne). Zakres badania zawężono do analizy predykatu mentalnego *wierzyć* w języku polskim w jego dwóch formach aspektowych (*wierzyć* i *uwierzyć*). Z korpusu PWN, który stanowi część Narodowego Korpusu Języka Polskiego pobrano wszystkie formy morfologiczne analizowanych czasowników. Z ponad 2300 przykładów wybrano formy osobowe czasowników tak aby możliwe było określenie cech podmiotu oraz zdania z dopełnieniem zdaniowym i frazowym, aby określić semantykę dopełnienia, co zredukowało liczbę analizowanych zdań do 815.

Analiza danych pokazała że tylko niektóre zmienne formalne wykazały różnice we współwystępowaniu ze zmienną semantyczną. W kodowaniu danych wyróżniono zatem następujące zmienne:

- semantyczną: semantyka dopełnienia⁴, która mogła przyjmować jedną z dwóch wartości: doświadczenie mentalne indywidualne lub doświadczenie mentalne kolektywne;
- formalne: aspekt, osoba, liczba czasownika oraz forma gramatyczna dopełnienia.

Zilustrujmy je przykładami: (s.135)

- (1) *Oczywiście nikt nie wierzy w koniec wojny w tym roku.* (doświadczenie mentalne kolektywne, aspekt niedokonany, 3 os., 1. poj., dopełnienie frazowe)
- (2) *Jestem z nim ponad 4 lata i uwierzcie mi, nie jest ideałem.* (doświadczenie mentalne indywidualne, aspekt dokonany, 2 os. (tryb rozkazujący), 1. mn., dopełnienie zdaniowe)
- (3) *Według jego relacji - dzień wcześniej, podczas kąpieli chłopczyk wypadł mu z ręki, z wysokości około jednego metra. (...) Lekarze nie uwierzyli jednak jego relacji.* (doświadczenie mentalne indywidualne, aspekt dokonany, 3 os., 1. mn., dopełnienie frazowe)

⁴ Jedynie w przypadku dopełnień zdaniowych i frazowych możliwe było określenie ich semantyki. W przypadku dopełnień zaimkowych nie było to możliwe.

- (4) *Osobiście wierzę, że nie jest to prawdopodobne.* (indywidualne doświadczenie mentalne, aspekt niedokonany, 1 os, 1. poj., dopełnienie zdaniowe)

W programie Excell kodowanie zdań przyjęło następującą formę:

	1	2	3	4	5	6	7	8
	Sentence	Source	Form	V_pers	V_Number	Obj_Form	Obj_Cog	Coder
1	Oto polska wieś, oto Polska cała w pigułce. Uwierzę w bezrobocie, biedę i powstanie dopiero wtedy, gdy na moim beskidzkim odludziu chętni do kopania ogródka lub przywiezienia drewna z lasu sami będą pukać do drzwi, a nie będę się musiał dopraszać o taką przysługę.	Typ tekstu: Prasa Tytuł: Wprost Nr: 11.07 Miejsce wydania: Poznań Rok: 1999	uwierzę	V1	VSg	ObjNP	Soc	MF
2	Wasył, 64-letni górnik, trzyma w rękach tornister. Prowadzi wnuczkę ze szkoły. - Uwierzy pan, że jeszcze pracuję w kopalni? - mówi. Ale teraz wożę już tylko górników windą w górę i w dół.	Typ tekstu: Prasa Tytuł: Rzeczpospolita Nr: 12.02 Miejsce wydania: Warszawa Rok: 2004	uwierzy	V3	VSg	ObjClause	Indiv	MF
3								

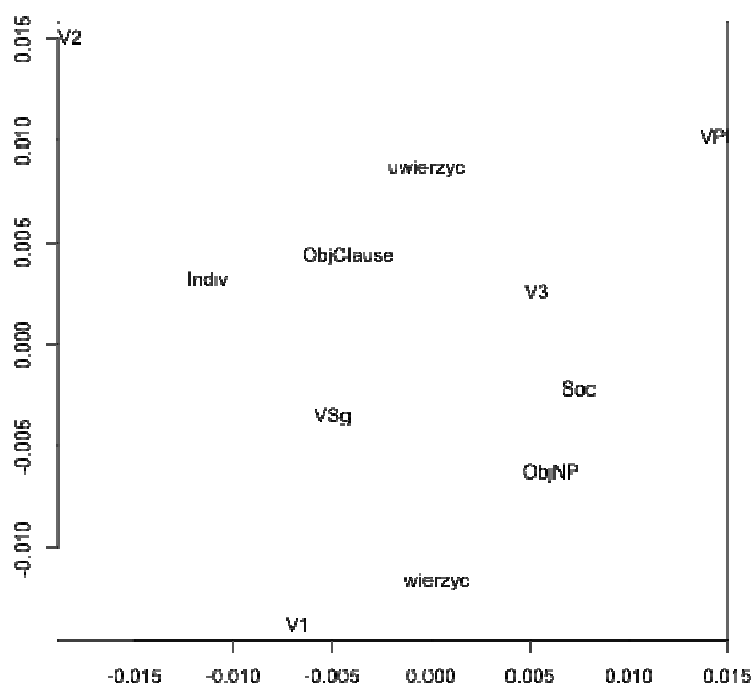
Rysunek 1 Fragment kodowania (anotacji) w arkuszu programu Excell.

Aby uzyskać wyższą sprawdzalność intersubiektywną, każdy przykład analizowany był przez dwóch językoznawców (stąd kolumna 8 z inicjałami osoby kodującej), w razie rozbieżności przykłady omawiano na spotkaniach zespołu badawczego. Jeśli nie udało się uzyskać konsensusu zdania takie usuwano z próby poddanej testom statystycznym.

Po przeanalizowaniu niemal tysiąca zdań możliwe było zawężenie ogólnego pytania badawczego do nieco bardziej szczegółowych hipotez badawczych:

- (I) Istnieje zależność między indywidualnym lub kolektywnym doświadczeniem umysłowym a osobą i liczbą czasownika (formą gramatyczną podmiotu).
- (II) Istnieje zależność między indywidualnym lub kolektywnym doświadczeniem umysłowym a formą gramatyczną dopełnienia.

Aby zbadać te intuicje wynikające z analizy zdań zastosowano eksploracyjną technikę statystyczną w formie analizy korespondencji wszystkich zmiennych. (s.136) Analiza korespondencji pozwala na zbadanie korelacji pomiędzy poszczególnymi zmiennymi, która zostaje przedstawiona w formie dwuwymiarowego wykresu. Wykres ten stanowi kompresję wielowymiarowych analiz frekwencji wszystkich zmiennych oraz relacji pomiędzy nimi. Korelacje pomiędzy zmiennymi wyrażone są poprzez odległości. Zmienne występujące blisko siebie na wykresie wykazują wyższy stopień korelacji niż te, które są od siebie oddalone. Analizę statystyczną przeprowadzono przy pomocy programu R. Reprezentuje ją wykres poniżej:



Wykres 1: Analiza korespondencji wszystkich zmiennych. Znaczenie skrótów wyrażeń na wykresie: Indiv – indywidualne doświadczenie mentalne, Soc – kolektywne doświadczenie mentalne; wierzyć – aspekt niedokonany, uwierzyć – aspekt dokonany; V1 – 1 osoba czasownika, V2 – 2 osoba czasownika, V3 – 3 osoba czasownika, VP1 – l. mnoga, VSg – liczba pojedyncza; ObjNP – dopełnienie frazowe, ObjClause – dopełnienie zdaniowe.

Analiza korelacji widocznych na Wykresie 1 pozwala na uściślenie hipotez badawczych do następującej postaci:

- (I) Indywidualne doświadczenie umysłowe znacząco częściej współwystępuje z dopełnieniem zdaniowym;
- (II) Indywidualne doświadczenie umysłowe znacząco częściej współwystępuje z podmiotami drugoosobowymi; (**s. 137**)
- (III) Indywidualne doświadczenie umysłowe znacząco częściej współwystępuje z czasownikiem dokonanym: *uwierzyć*;
- (IV) Kolektywne doświadczenie umysłowe współwystępuje częściej z dopełnieniem w formie frazy nominalnej;
- (V) Kolektywne doświadczenie umysłowe współwystępuje częściej z niedokonaną formą czasownika: *wierzyć*.

W związku z tym, że zmienna objaśniana – semantyka dopełnienia – przyjmuje wartości dychotomiczne (albo indywidualne doświadczenie umysłowe albo kolektywne doświadczenie umysłowe) do obliczenia prawdopodobieństwa sukcesu, a więc prawdopodobieństwa z jakim może wystąpić prognozowane współwystąpienie zmiennych zastosowano Regresję Logistyczną. Jej wyniki zaprezentowano w Tabeli 1 poniżej:

Tabela 1 Regresja Logistyczna dla indywidualnego doświadczenia mentalnego (N=324) i kolektywnego doświadczenia mentalnego (N=491) dla następujących zmiennych: aspekt, osoba, liczba i forma dopełnienia (za Fabiszak – Hebda – Konat 2012: 172).

Deviance Residuals:				
Min	1Q	Median	3Q	Max
-2.7199	-0.8518	0.3891	0.7276	1.8349
Coefficients:				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3510	0.2888	4.677	2.91e-06 ***
Aspectwierzyc	1.0410	0.1846	5.638	1.72e-08 ***
V_persV2	-0.4998	0.3228	-1.549	0.121
V_persV3	0.1510	0.1958	0.771	0.441
V_NumberVSg	-2.3292	0.2517	-9.254	< 2e-16 ***
Obj_FormObjNP	1.1309	0.1730	6.537	6.27e-11 ***
Null deviance: 1095.37 on 814 degrees of freedom				
Residual deviance: 880.83 on 809 degrees of freedom				
AIC: 892.83				
P	C	R2	Brier	
0	0.785	0.313	0.182	

Jak pokazują wyniki w Tabeli 1. Korelacja między Indywidualnym doświadczeniem umysłowym (oznaczone w kolumnie ‘Estimate’ przez ‘-’) a liczbą pojedynczą czasownika (i podmiotu) jest statystycznie znaczące ($p > .001$, oznaczone w kolumnie ‘Pr’ przez trzy gwiazdki ***). Jest to wynik, który przewidywało pierwsze sformułowanie hipotez badawczych, ale które pominięto w ich wersji uściślonej z powodu podobnej odległości na Wykresie 1 pomiędzy liczbą pojedynczą a obydwooma typami semantyki dopełnienia. Wyniki zaprezentowane w Tabeli 1 pozytywnie weryfikują uszczegółowione hipotezy (IV) i (V), które postulują współwystępowanie Kolektywnego doświadczenia mentalnego z frazą nominalną i z niedokonaną formą czasownika *wierzyć*. (s. 138)

Wyniki Regresji Logistycznej pozwalają nam stwierdzić, że konceptualizatorzy scen wyrażonych przez zdania z badanego korpusu częściej przypisują indywidualne doświadczenie mentalne *Experiensom* w l. poj. Poznanie kolektywne współwystępuje znacząco częściej z czasownikiem niedokonanym i dopełnieniem w postaci frazy rzeczownikowej.

Tezę o tym, że kolektywne doświadczenie mentalne częściej wyrażone jest za pomocą frazy nominalnej można odnieść do takich pojęć językoznawstwa kognitywnego jak skanowanie sumaryczne (ang. *summary scanning*) Langackera (np. 1991) oraz obiektyfikacja Szwedka (np. 2011). Wyniki analizy użycia czasownika *u/wierzyć* w języku polskim wskazują na to, że pojęcia czy wyobrażenia mentalne dostępne intersubiektywnie dużej grupie *Eksperiens*erów przyjmują formę nominalizacji będących rezultatem skanowania sumarycznego

lub obiektywizacji, gdyż szczegóły, na których wyrażenie pozwala dopełnienie zdaniowe, w przypadku wyobrażeń wspólnych (ang. *shared representation*) dla danej grupy komunikacyjnej - nie są potrzebne.

3. Korzyści i problemy wynikłe ze stosowania korpusów językowych w językoznawstwie kognitywnym

Wyżej opisane przykładowe badania są dobrą ilustracją korzyści, ale też problemów i ograniczeń związanych z zastosowaniem korpusów językowych w językoznawstwie kognitywnym. Wśród pożytków należy wymienić wyższy stopień intersubiektywności takich badań, natomiast wśród ograniczeń – zawężenie zakresu stosowalności twierdzeń językoznawstwa kognitywnego.

Sądzymy, że korpusy zwiększają stopień pewności, z jakim można przyjmować wyniki badań prowadzonych w językoznawstwie kognitywnym. Według metodologii nauki, stopień pewności z jakim można przyjmować wyniki badań zależy od stopnia w jakim spełnione są warunki intersubiektywnej komunikowalności badań. W naukach empirycznych, ważnym kryterium naukowości jest intersubiektywna komunikowalność i intersubiektywna sprawdzalność prowadzonych badań (Nowak 1977, Such 2006). Kmita (1976) formułuje te kryteria w taki sposób:

„Mianowicie zdanie proponowane na twierdzenie naukowe winno być (m. in.) tak sformułowane, aby mogło być poddane *kontroli intersubiektywnej*, w więc: (1) wykonalnej w zasadzie dla wszystkich reprezentantów odnośnej dyscypliny naukowej (2) przebiegającej w każdym indywidualnym przypadku analogicznie i prowadzącej, o ile nie zajdzie jakaś pomyłka, do analogicznych rezultatów”. (Kmita 1976: 112)

Kryterium to pozwala na uporządkowanie metod badawczych w lingwistyce. Naszym zdaniem stopień intersubiektywnej kontroli w przypadku analizy zdań stworzonych przez lingwistę jest mniejszy niż w przypadku analizy zdań zasłyszanych-(s. 139)

-nych. Względem analizy zdań zasłyszanych z kolei, oparcie badań o szczegółowo opisany korpus papierowy zwiększa stopień intersubiektywnej kontroli. Z racji otwartego dostępu oraz zestandaryzowanych procedur i narzędzi, elektroniczne korpusy językowe pozwalają na wielokrotne, powtarzalne analizy przeprowadzane przez kolejne grupy lingwistów. Tym samym spełniają zarysowane wyżej kryterium intersubiektywnej kontroli w większym jeszcze stopniu niż korpusy papierowe.

Nie znaczy to jednak, że metody te wzajemnie się wykluczają. Hipotezy wysnute na podstawie obserwacji i introspekcji lingwisty mogą zostać zamienione na pytania badawcze a te z kolei na hipotezy testowalne w ramach psycholingwistyki oraz językoznawstwa korpusowego (podobnie twierdzą Arppe, Gilquin, Glynn, Hilpert, Zeschel, 2010). W językoznawstwie kognitywnym pewne obserwacje poczynione przez Lakoffa czy Langackera na przykładach zasłyszanych czy wymyślonych stały się podstawą dla badań opartych o metody empiryczne. Jak zauważa Arppe et al., zbadanie problemu w ramach wielu różnych metodologii pomaga niwelować błędy i zbliża do adekwatnego ujęcia zjawiska (Arppe, Gilquin, Glynn, Hilpert, Zeschel, 2010, s. 3).

Wnioskowanie badacza oparte o korpusy językowe jest jednak również obciążone pewnym ryzykiem. Autorzy z nurtu językoznawstwa korpusowego zwracają uwagę na fakt, że w przypadku poszukiwania zbieżnych dowodów badacz musi pamiętać o jasnych procedurach wskazywania powiązań i interpretacji wyników. Niektóre relacje między pojęciami jakościowymi a ilościowymi nie są bowiem jeszcze dokładnie zbadane, jak na przykład relacja między częstością występowania a istotnością (Arppe, Gilquin, Glynn, Hilpert, Zeschel, 2010, s. 7).

Kolejne ograniczenie wiąże się z faktem, że badania korpusowe to obserwacja, nie eksperyment, ponieważ nie występuje w nich żadna zmienna kontrolowana (Such 2006). Stosując inne kryterium stawia linię podziału w tym samym miejscu Glynn, dzieląc dane lingwistyczne na dane znalezione (ang. *found data*) oraz dane wydobywane (ang. *elicited data*). Dane znalezione mają postać korpusu, wobec którego można następnie przeprowadzić analizę jakościową (wybranych zdań) albo ilościową. Dane wydobywane mają postać wyników eksperymentu psycholingwistycznego lub też odpowiedzi na pytania zadawane mówcy (Arppe, Gilquin, Glynn, Hilpert, Zeschel, 2010, s. 7). Jeśli zatem korpusy przynależą do sfery obserwacji, powinien być na nie nałożony wymóg reprezentatywności. Niestety korpusy językowe nie mogą być reprezentatywne, w ścisłym sensie reprezentatywności przyjmowanej przez nauki społeczne:

„Uznamy próbę za reprezentatywną dla populacji, z której została dobrana, jeśli zagregowane cechy tej próby ściśle odzwierciedlają te same zagregowane cechy populacji.”
(Babbie, 2006, s. 210).

Ponieważ nie są znane wartości natężenia danych cech w „całości języka” (jakkolwiek może ona być rozumiana), spełnienie powyższego warunku w ścisłym sensie nie jest możliwe. Zdarza się, że językoznawcy określają korpusy (s. 140) mianem „reprezentatywnych”, „równoważonych” lub „wiarygodnych” (np. Leech, 2006; Mautner, 2009). Takie określenia mają jednak charakter jedynie przybliżeń. Dlatego właśnie, hipotezy testowane na korpusie nie podlegają falsyfikacji w ścisłym sensie. Przecież fakt, że jakiegoś zjawiska nie stwierdzono w korpusie nie może być przesłanką do stwierdzenia, że nie występuje ono wcale. Podobnie, stwierdzenie z pewnością występowania jakiegoś zjawiska w korpusie nie świadczy o tym, że z taką samą pewnością można twierdzić o występowaniu tego zjawiska w całości wypowiedzi językowych. Jednak potwierdzenie prawa dla pewnej ilości przypadków wzmacnia przypuszczenie, że prawo to obowiązuje także dla pozostałej części zbioru (Such 2006, s. 60). Na tej podstawie można stwierdzić, że w ramach kognitywnego językoznawstwa korpusowego tworzy się twierdzenia numerycznie ogólne: orzekające coś o całości zbioru, którego wszystkie elementy są znane.

Podsumowując można stwierdzić, że zastosowanie korpusów językowych w badaniach językoznawstwa kognitywnego przyczynia się do zwiększenia stopnia intersubiektywności i pewności wyników tych badań. Jednocześnie dalekosiężne cele, jakie stawiało sobie językoznawstwo kognitywne, ulegają

zawężeniu, gdyż wyniki badań korpusowych są relatywizowane w swej stosowalności do wielkości i jakości badanego korpusu.

Bibliografia:

- Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., Zeschel, A. 2010. Cognitive Corpus Linguistics: Five points of debate on current theory and methodology. *Corpora*, 5(1),
- Babbie, E. R. 2006. *Badania społeczne w praktyce*. Warszawa: Wydawnictwo Naukowe PWN.
- Bernárdez, E. 2007. Synergy in the construction of meaning. (in:) M. Fabiszak (Red.)
- Cameron, L. 2010. *Metaphor and Reconciliation: The Discourse Dynamics of Empathy in Post-Conflict Conversations*. Routledge Studies in Linguistics. New York, USA: Routledge.
- Evans, V., Green, M., 2006. *Cognitive Linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Fabiszak, M. (Red.). 2007. *Language and Meaning. Cognitive and Functional Perspectives*. Frankfurt am Mein: Peter Lang.
- Fabiszak, M., Hebda, A., 2012. O metaforze we wczesno-nowo-angielskich medycznych tekstach specjalistycznych, popularnonaukowych i dialogach. (w:) P. Nowakowski, K. Stroński, M. Szczyszek (Red.). Wspólnoty Komunikacyjne. Poznan: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk.
- Fabiszak, M., Hebda, A., Konat, B. 2012. Dichotomy between private and public experience: The case of Polish wierzyć 'believe'. (w:) *Online Proceedings of UK-CLA Meetings*, 1 (1).
- Fauconnier, G. 1994. *Mental Spaces: Aspects of Meaning Construction in Natural Language* (2. wyd.). Cambridge University Press.
- Fillmore, Ch. 1975. *An alternative to checklist theories of meaning*. Proceedings of the Berkeley Linguistics Society.
- Fillmore, C. J. 2006. Frame semantics. [w:] D. Geeraerts (Red.).
- Geeraerts, D. 2006. (Red.), *Cognitive linguistics: basic readings*. Walter de Gruyter.
- Glynn, D., Fischer, K. 2010. (Red.) *Corpus-Driven Cognitive Semantics. Quantitative approaches*. Cognitive Linguistics Research. Berlin: Mouton de Gruyter.
- Glynn, D., Robinson, J. W przygotowaniu. (Red.) *Polysemy and Synonymy. Corpus methods in Cognitive Linguistics*. Human Cognition. Amsterdam: John Benjamins.
- Goldberg, A. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press, USA.
- Grady, Joseph E. 1997. Foundations of meaning: Primary metaphors and primary scenes. Ph. D. dissertation at the University of Berkeley.
- Gries, S. T. 2007. Introduction. [w:] A. Stefanowitsch, S. T. Gries (Red.).
- Hundt, M., Nesselhauf, N., Biewer, C., 2006. (Red.), *Corpus Linguistics and the Web (Language & Computers 59)*. Editions Rodopi BV.

- Johnson, Ch. 1999. Constructional grounding: The role of interpretational overlap in lexical and constructional acquisition. Ph.D. dissertation, University of California, Berkeley.
- Johnson, M. 1987. *Body in the Mind*. Chicago: Chicago University Press.
- Lakoff, G. 1991. Cognitive versus generative linguistics: how commitments influence results. *Language & Communication*, (11), 53–62.
- Lakoff, G. 1987. *Fire, Women and Dangerous Things. What categories can reveal about the mind*. Chicago: Chicago University Press.
- Lakoff, G. 1999. Cognitive Models and Prototype Theory. [w:] E. Margolis, S. Laurence (Red.).
- Lakoff, G., Johnson, M. 1999. *Philosophy in the Flesh : The Embodied Mind and Its Challenge to Western Thought*. Basic Books.
- Lakoff, G., Johnson, M. [1980] 2003. *Metaphors We Live By* (2 wyd.). University Of Chicago Press.
- Langacker Ronald. 1991. *Concept Image Symbol: The Cognitive Basis of Grammar*. The Hague/Berlin: Mouton de Gruyter.
- Langacker, R. 1999. *The Foundations of Cognitive Grammar: Volume I: Theoretical Prerequisites*. Stanford University Press.
- Langacker, R. W. 2009. *Gramatyka Kognitywna: Wprowadzenie*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych Universitas.
- Leech, G. 2006. New Resources, or Just Better Old Ones? The Holy Grail of Representativeness. [w:] M. Hundt, N. Nesselhauf, C. Biewer (Red.).
- Lewandowska-Tomaszczyk, B. 2002. Polysemy research: methodology and mechanisms. [w:] B. Lewandowska-Tomaszczyk, K. Turewicz [Red.]. *Cognitive Linguistics Today*. Frankfurt: Peter Lang. 81-96.
- Lewandowska-Tomaszczyk, B. 2005. *Podstawy językoznawstwa korpusowego*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Lewandowska-Tomaszczyk, B., 2006. Metody empiryczne i korpusowe w językoznawstwie kognitywnym. (w:) P. Stalmaszczyk (Red.)
- Lewandowska-Tomaszczyk, B., Turewicz. K. [Red.]. *Cognitive Linguistics Today*. Frankfurt: Peter Lang.
- Margolis E., Laurence S., (Red.), *Concepts: core readings*. MIT Press.
- Mautner, G. 2009. Checks and Balances: How Corpus Linguistics can Contribute to CDA. [w:] R. Wodak, M. Meyer (Red.)
- Nerlich B., Hamilton C.A., Rowe V., 2002. Conceptualising foot and mouth disease: The socio-cultural role of metaphors, frames and narratives. *metaphorik.de* (02): 90-108.
- Nowak, L. 1977. *Wstęp do Idealizacyjnej Teorii Nauki*. Wybrane Zagadnienia Metodologii Nauk, Naukoznawstwa i Informacji Naukowej. Warszawa: Państwowe Wydaw. Naukowe.
- Nowakowski, P., Stroński, K., Szczyszek, M. (Red.). *Wspólnoty Komunikacyjne*. Poznań: Wydawnictwo Poznańskiego Towarzystwa Przyjaciół Nauk.
- Popper, K. R. [1936] 2002. *Logika Odkrycia Naukowego* (Wyd. 2.). Warszawa: Wydawnictwo Naukowe PWN.
- Semino, E. 2011. The adaptation of metaphors across genres. *Review of Cognitive Linguistics* (9, 1): 130-152.

- Stalmaszczyk, P. 2006. (Red.) *Metodologia językoznawstwa. Podstawy teoretyczne*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Stefanowitsch, A., Gries S. T., 2007. (Red.), *Corpora in Cognitive Linguistics: Corpus-Based Approaches to Syntax and Lexis*. Walter de Gruyter.
- Szwedek A. 2011. The Ultimate source domain. *Review of Cognitive Linguistics* (9.2): 341-366.
- Taylor, J. 2007. *Gramatyka kognitywna*. Kraków: UNIVERSITAS.
- Tuggy, David. 1993. Ambiguity, polysemy and vagueness. *Cognitive linguistics* (4-3): 273-290.
- Tuggy, David. 1999. Linguistic evidence for polysemy in the mind: A response to William Croft and Dominiek Sandra. *Cognitive Linguistics* (10-4): 343-368.
- Wodak, R., Meyer M. (Red.), *Methods of Critical Discourse Analysis* (Second Edition.). Sage Publications Ltd.
- Zlatev, J., Racine, T., Sinha C., Itkonen, E., 2008. (Red.) *The Shared Mind: Perspectives on intersubjectivity*. Amsterdam: John Benjamins.